formed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system modules and components in the embodiments described above should not be understood as requiring such separation in all embodiments, and it should be understood that the described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

Particular embodiments of the subject matter have been described. Other embodiments are within the scope of the following claims. For example, the actions recited in the claims can be performed in a different order and still achieve desirable results. As one example, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In some cases, multitasking and parallel processing may be advantageous.

What is claimed is:

1. A method comprising:

receiving training data for training a machine learning model to perform a particular machine learning task, the training data comprising a plurality of training inputs to the machine learning model;

determining a plurality of data augmentation policies, wherein each data augmentation policy has a plurality of data augmentation policy parameters that define a procedure for transforming training inputs before the training inputs are used to train the machine learning model, the determining comprising, at each of multiple time steps:

generating a current data augmentation policy based on, for one or more previous time steps, a quality measure of a data augmentation policy generated at the previous time step, wherein for each previous time step:

the quality measure of the data augmentation policy generated at the previous time step characterizes a performance measure of the machine learning model on a set of validation data comprising a plurality of validation inputs to the machine learning model after the machine learning model has been trained on the training data using the data augmentation policy generated at the previous time step,

wherein the machine learning model has not been trained on the validation data;

training the machine learning model on the training data using the current data augmentation policy, comprising:

selecting a batch of training inputs from the training data;

determining an augmented batch of training inputs by transforming the training inputs in the batch of training inputs in accordance with the current data augmentation policy; and

adjusting current values of parameters of the machine learning model based on the augmented batch of training inputs; and

determining the quality measure of the current data augmentation policy using the machine learning model after the machine learning model has been trained on the training data using the current data augmentation policy;

selecting a final data augmentation policy based on the quality measures of the determined data augmentation policies; and

generating a final trained machine learning model by training a final machine learning model on the training data using the final data augmentation policy.

2. The method of claim **1**, wherein the particular machine learning task is an image processing task comprising classification or regression.

3. The method of claim **1**, wherein:

each data augmentation policy comprises one or more sub-policies;

each sub-policy comprises a sequence of one or more transformation tuples, wherein for each transformation tuple, the data augmentation policy parameters define: (i) a transformation operation, and (ii) a magnitude of the transformation operation; and

transforming the training inputs in the batch of training inputs in accordance with the current data augmentation policy comprises, for each training input:

identifying a sub-policy included in the data augmentation policy; and

transforming the training input in accordance with the identified sub-policy by sequentially applying each transformation tuple included in the identified sub-policy to the training input.

4. The method of claim **3**, wherein identifying a sub-policy included in the data augmentation policy for the training input comprises randomly sampling a sub-policy included in the data augmentation policy.

5. The method of claim **3**, wherein applying a transformation tuple to the training input comprises:

applying the transformation operation from the transformation tuple with the transformation operation magnitude from the transformation tuple to the training input.

6. The method of claim **3**, wherein:

for each transformation tuple, the data augmentation policy parameters further define a probability of applying the transformation operation; and

applying a transformation tuple to the training input comprises:

applying the transformation operation from the transformation tuple with the transformation operation magnitude from the transformation tuple to the training input with the transformation probability from the transformation tuple.

7. The method of claim **1**, wherein the machine learning model is a neural network, and adjusting the current values of the machine learning model parameters based on the augmented batch of training inputs comprises:

determining a gradient of a loss function using the augmented batch of training data; and

adjusting the current values of the machine learning model parameters using the gradient.

8. The method of claim **1**, wherein:

generating the current data augmentation policy based on, for one or more previous time steps, the quality measure of the data augmentation policy generated at the previous time step comprises generating the current data augmentation policy using a policy neural network in accordance with current values of policy neural network parameters; and

the policy neural network is trained by reinforcement learning techniques based on a reinforcement learning reward signal, and at each time step, the reinforcement